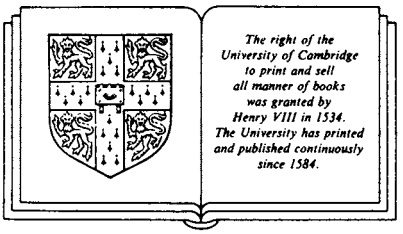
The Cunning of Reason

Martin Hollis



Cambridge University Press

Cambridge
New York Port Chester
Melbourne Sydney

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1987
Reprinted 1989

British Library cataloguing in publication data

Hollis, Martin
The cunning of reason.
1. Reasoning (Psychology)
I. Title
128'.3 BF441

Library of Congress cataloguing in publication data

Hollis, Martin
The cunning of reason.
Bibliography.
Includes index.
1. Reason. I. Title.
BC177.H65 1988 128'.3 87–14630

ISBN 0 521 24879 5 hard covers
ISBN 0 521 27039 1 paperback

Transferred to digital printing 2003

RB

Contents

I

The sovereign artificer

God speaks to Adam:

> I have given you, Adam, neither a predetermined place nor a particular aspect nor any special prerogatives in order that you may take and possess these through your own decision and choice. The limitations on the nature of other creatures are contained within my prescribed laws. You shall determine your own nature without constraint from any barrier, by means of the freedom to whose power I have entrusted you. I have placed you at the centre of the world so that from that point you might see better what is in the world. I have made you neither heavenly nor earthly, neither mortal nor immortal, so that like a free and sovereign artificer you might mould and fashion yourself into that form you yourself shall have chosen.

These ringing sentences come from the the *Oration on the Dignity of Man* which Pico della Mirandola wrote in 1484. They capture a Renaissance humanist vision of man as special in the cosmos, and should catch recent social theorists in two minds. On the one hand God places Adam at the centre of the world and makes him his own artificer. That is just right for an individualism grounded on the idea that man is an active subject, who fashions the fabric of the social world. On the other hand God prescribes no constraints from any barrier. That seems to put Adam beyond the pale of every kind of science. However tempting active subjects may be to political theories of liberty, science needs objects.

Those last three words would be decisive if social theorists were still mesmerised by Newtonian mechanics and nineteenth-

century physics. The social sciences took shape in an age divided from Pico's by the scientific revolution. They inherited a Newtonian picture of nature as a mechanical, law-governed, humanly empty system, whose workings were hidden from the naked eye, rather as the springs and wheels of a fob watch are invisible to those who see only its face. Natural science had triumphed by learning to prise the back off the watch. It had found the forces acting on the cogs and learnt the secrets of the innermost particles of matter. Could the same not be done for human nature and human society?

This question still itched in the 1950s, when I was first exposed to it. By then the prevailing philosophy of science was empiricist, more concerned with the success of prediction and less with structures beyond the reach of experience. Philosophers were thinking of causal laws in terms more of interesting correlations than of necessitating forces. David Hume, quite newly promoted to the pantheon, was responsible for that, and social scientists, although divided about the merits of his analysis of causation, found nothing subversive in his statement of their task. He had published *A Treatise of Human Nature* in 1739 as a foundation for 'a complete system of the sciences'. It is evident, he had said in the introduction, that all sciences have a relation, greater or less, to human nature. 'Even Mathematics, Natural Philosophy and Natural Religion are in some measure dependent on the science of MAN; since they lie under the cognisance of men, and are judged by their powers and faculties.' Hence a science of man, grounded in an empirical study of 'Logic, Morals, Criticism and Politics' would comprehend almost everything which can tend to the improvement of the human mind. Its method was to be 'experience and observation', applied to 'men's behaviour in company, in affairs and in their pleasures'. Its aim was 'explaining all effects from the simplest and fewest causes'.

We reach 1987 without anything at all like a complete system of the sciences. Indeed there is no longer an agreed natural science blueprint by which to test the character of social science. Thanks to Popper, Quine, Kuhn, Feyerabend, Lakatos and others, the thought that science lies under the cognisance of man has undercut common assumptions about scientific objectivity and the independence of facts. Where there used to be foundations for

scientific knowledge in unvarnished observation and a hypothetico-deductive procedure for testing theories, it is now argued that there is no 'unvarnished news' (Quine's evocative phrase) and that no observation or test can be innocent of theory. Natural scientists may clock into their laboratories undismayed, but social scientists have never worked under laboratory conditions. If the interpretation of social facts is always the interpretation of interpretations, then the hermeneutic aspects of social science become dizzying.

That is a reference to the hermeneutic or interpretative tradition in social thought, which descends from German idealist reflections on the writing of history. It was never cowed by the insistence that social sciences must be *sciences*, and is now becoming urgent. Its leading category is Meaning, characterised by Wilhelm Dilthey as 'the category which is peculiar to life and to the historical world'. Human life, he said, can be understood only by means of categories which do not apply to our knowledge of physical reality. Meaning, value, purpose, development and ideal are such categories. They all depend on the fact that the connectedness of a life can only be understood through the meaning that individual parts have for understanding the whole. 'Life does not mean anything other than itself. There is nothing in it which points to a meaning beyond it.'[1]

The guiding imperative here is that action has meanings which can be understood only from within. If we ask why the historian, as typical student of mankind, differs from the physicist, as typical student of nature, four instructive thoughts about meaning suggest themselves. (I shall put them in general terms of my own, since I cited Dilthey more out of respect for a founding father than because his actual words remain in circulation.) Firstly *experience* has meaning for people. They find in it signs of order, both natural, as when they classify animals or systematise the movements of planets, and human, as when they respond to a gesture or celebrate the marriage of a friend. They read their experience not only scientifically but also aesthetically and morally. They extract predictions, lessons and ideals from it. They impose themselves

[1] Dilthey (1926), vol. vii, p. 224. I have followed the convention of referring to works by citing the year of publication of a standard edition in brackets, e.g. (1926), even when, as here, it is not the year of origin.

and their concerns on it too, and do so in ways which vary between individuals, groups and cultures. An atomic particle is passive by contrast. If it has experience, it kindly refrains from attaching subjective meaning to it. It has no beliefs about the world which need worry a physicist.

Secondly *utterances* have meaning for people. Words and sentences are not just physical disturbances but communications in accordance with rules for their use. They are read and answered in a language governed at least largely by convention. At the same time speakers have both intention and motive in using them. The speaker's meaning and the utterance's meaning can go adrift in ways foreign to objects in nature. Here lie hermeneutic circles of the most daunting kind, which do not trouble the physicist.

Thirdly *actions* have both instrumental and expressive meaning. Both adjectives involve several kinds of meaning, depending on the context of enquiry: what the actor did, intended or aimed to achieve, what the action signified or symbolised and how it was embedded in a network of values or relations are some of them. Since this book is about understanding actors and actions, I shall not linger now. Suffice it to say that political manoeuvre, sexual courtship and religious observance (to pick almost at random) mix intention, motive, convention and propriety in various ways. There is a sort of Bermuda triangle, marked by questions of signs and symbols, of norms and principles and of means and ends, which swallows many a student of mankind. Physicists lose no sleep over it.

Fourthly *ideas* have meaning. People's reading of their situation alters with changes in their beliefs and expectations, as later chapters will insist. Among the ideas which move them are ideas about what moves them. Social theory, being itself in circulation among its subjects, is tied to its own tail. These are matters of meaning in various senses of that nimble term. That molecules have no thoughts about molecules must be a great relief to the physicist.

No quick conclusions follow from these differences between history and physics, but there is enough to pause for. It is not just that the historian needs an insider's view when reconstructing particular episodes in the lives of singular individuals. The suggestion is that an insider view is needed for every study of social life, because the social *is* what it *means*. Because life means noth-

ing beyond itself, a peculiar kind of understanding is called for. This is plausible only if made very sweeping. Any fragment of social life certainly 'means' something beyond itself. The Russian Revolution, for instance, was not a self-contained episode, and especially not one which could be understood just by knowing what the actors had in their conscious minds. Even broad periods of history or complete cultures are set in a natural world and peopled by bodily creatures, as the study of technology or medicine makes abundantly clear. So there is no meaning beyond life itself only if some grand idealist proposition is offered, internalising reality to consciousness and perhaps finding an overall pattern and movement there, too global for individual minds to grasp in their particular time and place.

It will sound as if I am preparing to live up to my Hegelian title by conjuring up a World Spirit to direct History. The Cunning of Reason is Hegel's term for a hidden dynamic or dialectic which sums the consequences of actions in ways unforeseen by the actors. But I have no such grand ambition, witness the outline of the book which I shall give in a moment. I sound the bold idealist alpenhorn only to introduce the tensions which arise if one accepts that meanings matter but denies that nothing else does. The social sciences have always been caught between history and physics (to use a brisk shorthand), and the recent intrusion of hermeneutics into the discussion of objectivity in natural science has not relieved the tensions.

Natural science is happy to take a spectator's or outsider's view of the workings of nature. Explanation naturally takes the form of finding causes for effects, which are systematically linked by causal laws. Thinking in terms of causes, which produce or necessitate their effects and make them inevitable, would do nicely, were it not for epistemologists complaining that we cannot know of such connections. But acceptance of a softer Humean determinism, which (roughly) makes nature regular rather than forceful, does not undermine the idea that objectivity is for spectators. Nor does acceptance that there may be a random streak in nature, which makes complete predictability impossible. The suggestion from Quine and others that the mind can never be wholly self-effacing in what it regards as knowledge of nature is more worrying. But any retreat from a spectator view

will be grudging, since (odd cases like psycho-somatic illnesses aside) we still believe deeply that the workings of nature are a test of belief, not an effect of it.

An outsider's view of social life is more precarious. The urge to achieve one in the name of science has produced some uncompromising approaches, like behaviourism, which try to dispense with actors' inner states. But I presume that the actors' desires and beliefs can be allowed to matter for scientific explanation, provided that they have causes and effects. This line avoids direct conflict with cherished beliefs about free will. Indeed, if free action is action which gets the agent what he wants and is done because the agent had so calculated, freedom seems to presuppose a deterministic system of cause and effect. The actors themselves are not always the best judges of what they are doing, why and with what likely consequences. So the social sciences can increase our freedom by explaining us to ourselves and showing us how to achieve more of what we want. Thus microeconomics can plausibly be seen as a theory of rational choice which systematises 'insider' data in conformity with the demands of causal explanation, and hands the results back to the insiders as an aid to better decisions. Why I call this neat view precarious will emerge later.

By contrast, a hermeneutic approach to social life gives primacy to the world from within. It too can be uncompromising. 'The central concepts which belong to our understanding of social life are incompatible with concepts central to the activity of scientific prediction', declared Peter Winch in *The Idea of a Social Science* (1958, p. 94), his robust, well-argued Wittgensteinian assault on causal thinking about action. But in so far as free will is the issue, one should not assume that what a hard-nosed outsider approach denies is safe in hermeneutic hands. If the meaning of action is exhausted by the rules and cultural forms of social life, then Adam is no more the maker of his own history than the atom is of its own explosions. Yet there needs to be more to the meaning of action than what a particular agent took himself to be up to on specific occasions. Otherwise 'understanding' is merely describing or redescribing, and serves at most to set the explanatory agenda. Adam's claim to be a sovereign artificer seems to depend on Meaning being more than the alpha and less than the omega of understanding.

That ushers in some awkward epistemological questions. The

hermeneutic imperative is to understand action from within. How do we know when we have succeeded? How can we tell a better interpretation from a worse one? The answer cannot be meekly to accept the actors' own verdict. For, firstly, there is a notorious Other Minds problem in understanding the actors' own understanding, and it is further complicated if the actors are to be the judges of its solution. This is plainest in anthropology but Other Cultures are just a tuppence-coloured version of Other Minds. Secondly, more goes on in the social world than the actors, singly or collectively, notice and understand. Actors' accounts can be not just incomplete but also mistaken. Many actors are well aware of it – otherwise there would be no research funds for social science.

So we need a vantage point for assessing insider accounts, and one which is epistemologically robust. We need a category which lets the enquirer know when he has identified what the actors are doing and when he has understood why they are doing it. It must respect Adam's sovereign artifice and yet allow a wider view. That sets a riddle: 'When is a science not a science?' The best answer in hermeneutic vein seems to me: 'When it studies rational action'. At any rate, that is the answer to be explored in this book. The category which lets us make most objective yet interpretative sense of social life is Rationality.

The thought is not novel. It occurred notably to Max Weber, whose comments at the beginning of *Economy and Society* provide the best starting point. 'The science of society', he wrote, 'attempts the interpretative understanding of social action.' In action is included 'all human action, when and in so far as the acting individual attaches subjective meaning to it'. By social action is meant action 'which takes account of the behaviour of others and is thereby oriented in its course'. Subjective meanings are not simply private, but are connected with a public system of values. Social action, however, is not just conforming action, since individuals can have goals of their own and calculate the means to them. The proportion of calculation to conformity varies with the form of society, especially across the divide between traditional societies and today's emerging rational–legal Westernised societies. But the key to understanding is to find the rational element in what social agents do.

The theme is not altogether clear, because Weber then lists four

types of action, only two of which are labelled rational, and adds that most actions are of more than one type. The four are the instrumentally rational (*zweckrational*), the expressively rational (*wertrational*), the traditional and the affective. Of these, the idea of *zweckrational* action is readily grasped, being the familiar economists' notion of action which embodies the choice of the likeliest means to a given end. Understanding it, however, is not merely a matter of identifying the agent's ends (by reproducing his preferences) and the calculation which resulted in the choice (by reproducing his beliefs and information processing). Weber proposes the construction of an 'ideal-type' of the kind found in microeconomic theory, which lays out an optimal solution to the agent's problem of maximising the value of a variable subject to constraints. The ideal-type serves as a yardstick. Where the agent scores full marks, his action is thereby understood. Where the agent scores less than full marks, his departures need a further explanation, but the ideal-type has done the crucial job of identifying what they are. This method extends to all instrumental choices of action, even if we have no fully articulated theory. For instance the first step to understanding a general's decision in a battle is to ask what the optimal decision would have been. I shall say no more now about the merits and snags of an 'economic' notion of rationality, since they occupy the next chapters of the book.

Action is *wertrational* when the goal is so dominant for the actor that it drives out calculation of means and consequences. Acts of heroism and self-sacrifice are examples and, more broadly, some acts done from duty or principle. Understanding is a matter of grasping why the values expressed had such overriding significance for the actor. This is much less clear, especially since Weber distinguishes *wertrational* from traditional action, which might seem to be its obvious case. Traditional action is defined as 'the expression of settled custom' and glossed as often 'simply a dull reaction to accustomed stimuli'. Is conforming to a norm because it is the norm to be thought of as '*wertrational*' or as 'traditional'? The difference, if any, is further obscured by Weber's remarking that most actions are of mixed types (including the 'affective' type, where the agent is prompted by a simple, unreflective desire). Presumably there is a distinction here between principled action in a largely rational–legal society and norm-governed

action in a largely pre-modern one. For the moment, however, I note that there appears to be a second, expressive kind of rational action, whose character will become urgently puzzling as the book goes along.

Obscurities aside, Weber offers to put teeth into hermeneutics by having the social scientist first find the actor's own reasons for action and then judge their adequacy. This seems to me an intensely plausible reading of how microeconomics contributes to social understanding and how, therefore, the category of Meaning can be usefully glossed in terms of Rationality. At the same time, however, he also insists that explanation must be adequate both at the level of meaning and at the causal level. The former requires a method of understanding (*Verstehen*), which starts with 'empathy', or an awareness of what a person is doing akin to perception, and progresses to rational reconstruction through ideal-types. The latter calls for something Humean – causal explanation (*Erklären*) in the form of social statistics to show that what has been found by understanding is also typical of a common type of behaviour.

That sets a puzzle about the relation of the two levels. The idea, I think, is that rational reconstruction yields merely possible explanations, which become reliable only if they have empirical, predictive support. At any rate one common complaint about hermeneutic approaches is that they knit possible caps which fit the actors' heads but provide no way of knowing which cap the actors were actually wearing. Yet, if the final test of an interpretation is at the causal level, then *Verstehen* is ultimately only a heuristic device, a short cut to explanations which are finally in other terms. That would be consistent with the thinking of 'positive' social science, where conjectures and heuristics are welcome from all sources, provided that the logic of validation remains one of predictions derived from causal generalisations and upheld by experience. Weber himself seems to regard *Verstehen* sometimes as a distinct and genuine form of explanation, needing only reassurance at the causal level, and sometimes as an aid to causal explanation. He thus leaves unresolved the deeper puzzle of how an insider's or agent's point of view finally relates to an outsider's or spectator's one for purposes of social science.

I am happy to leave this ambiguity to the end of the book.

Meanwhile these hasty references to Weber have served to set the agenda. Revert to God's promise to make Adam his own free and sovereign artificer, who takes his place, aspect and prerogatives by his own decision. I shall be asking whether this humanism offers terms which a social science can accept by modelling Adam as a rational agent. The challenge is not just to be able to account for individual action within given parameters. 'Sovereign artificers' also make their places and prerogatives – the social fabric itself. Luckily 'Adam' is ambiguously singular or plural. What is not intentionally chosen by individuals singly can be jointly chosen. What is not chosen even in concert may yet be the intelligible sum of the consequences of individual choices. If the idea works, Adam can start the book as an individual rational agent and end it as mankind at large. 'Rational' is the key mystery term, to be decoded as the problems of understanding social action emerge.

With that hint at the theme, I turn now to a sketch of its components, using simple, clean lines to give an aerial view of the stages of the argument. Aerial views can be very misleading about the terrain at ground level. Expert readers will be aware that it is tangled, full of thorn bushes and occupied by argumentative tribes. Others may be lulled into the false security which comes from being handed a marked map in advance of trying the terrain for themselves. They may find that the outline makes better, because truer, sense after the expedition. But, although the clarity of the next few pages is only superficial, I hope that a marked map will give a helpful sense of direction.

There is no mystery about what the term 'rational' means to start with. Chapter 2 sets out a standard anatomy of rational choice by a single individual, intent on maximising his net utility within constraints. The anatomy is borrowed from microeconomics in a form which leads easily into the theory of *n*-person games. In chapter 3 Eve joins Adam, and it is shown how what is a parameter for each singly can be within their power to fashion collectively. Microeconomics is implicitly a very general theory of social action and even of the social framework. I shall treat *homo economicus* as an ambitious conjecture about *homo sapiens*, which applies wherever people make instrumentally rational

choices. Game-theoretic analysis of co-operation and competition not only crosses the apparent divide between the economic and the political or social but also offers to exhibit norms and institutions as deposits from previous games. We have the heady prospect of an individualist, contractarian theory of the social world.

This vision is not remotely plausible if the analysis is confined to the intended and desired consequences of individual choices. Some games solve co-ordination problems. For instance the British convention of driving on the left can be so construed. Other games end with a collective sum of individually rational choices which no one wants, although there is another possible outcome which all would have preferred. Here lies the notorious Prisoner's Dilemma, whose importance is clearly seen if one reads it into Hobbes' *Leviathan* as his account of why 'covenants without the sword are but vain breath'. In deference to Hobbes I term the crux the 'Leviathan trap' and threaten to bring contractarian thinking to an impasse. This is where the Cunning of Reason comes in. Hegel was not the first to spot that the consequences of individual actions can sum to unintended yet systematic outcomes. But his phrase catches my imagination. I do not respond to his suggestion that a World Spirit may be weaving a pattern in history in some amiable, or at least progressive, way. The attraction lies partly with 'cunning' and partly with 'reason'. The cunning springs surprises, which can explain collective mischief produced by individual rationality as well as public good produced by individual self-interest. 'Reason' suits the idea that an apparent snag of rational-choice theory can be transformed into a powerful extension of the theory.

The thought is trailed in chapter 4, where the Cunning of Reason makes the first of three appearances in the book. But I conclude that it cannot be incorporated until we know more about the motivation of a rational agent. Chapter 5 uncovers the Humean philosophy of mind involved in assuming that only desires can motivate and that a rational agent's desires need only be consistent. Taken seriously, these assumptions yield absurdities so large that chapter 6 substitutes a Kantian philosophy of mind for the Humean one. The ideal-type rational agent has good

reasons for his desires and beliefs and hence for his actions. He is likened to a Kantian moral agent. This makes it easier to understand how it can be rational to act on principles, and it gives leverage on the jaws of the Leviathan trap.

Having peopled the ideal-type world with agents who have rational preferences, I turn in chapter 7 to a new puzzle about rational belief. A rational agent needs to know the consequences, or likely consequences, of his options. But what if they depend on what other agents will decide to do, which depends on what they expect others to decide? A radical indeterminacy enters, rather as if tomorrow's weather responded to today's forecasts. Expectations are importantly generative, I argue, and cannot be dealt with by including rational expectations in the rational agent's stock of information. Decision infects prediction in a way alien to natural science and unsettling for ideas of probability.

This radical indeterminacy makes nonsense of the initial assumption that rational agents are maximisers.' Prompted by Herbert Simon's theory of 'bounded' rationality, chapter 8 makes Adam a 'satisficer', whose search for information is only as complete as it is rational for it to be. This move also leads us into some features of organistions. It becomes urgent that Organisation Man can have radically plural goals and hence 'satisfices' for the deeper reason that he must satisfy several incommensurable claims upon him. The attempt to analyse organisations as if they were individuals starts to go into reverse, with the suggestion that individuals are like organisations. A new Adam emerges, one whose reasons for action derive, partly but importantly, from his social positions and roles. *Homo sociologicus* takes the stage.

That will not be too dramatic a revision, if the Cunning of Reason can do more to organise the social world for *homo economicus*. Chapter 9 takes the business of unintended consequences a step further. Perhaps individual choices, considered collectively, can generate social structures which have their own systematic consequences. If so, *homo sociologicus* can still be *homo economicus* socialised. But functionalist notions of constraints, which serve the dynamics of the social structure, are too strong for Adam's comfort, if he is to survive as a sovereign artificer. There is more promise in the idea that systems consist of rules, which not only

constrain but also enable. That sets a difficult question about the relation of rules to choices or, as chapter 10 puts it, of roles to reasons. Adam is there presented as a rational role-player, who has 'distance' within each role and room for manoeuvre between them. Social life is a game created and played by actors, who construct their future from their available past.

It is then time to reopen the question of *Verstehen* and *Erklären*. Chapter 11 resumes the contrast and finds trouble with both halves. It offers to make honest epistemology of a hermeneutic approach by deploying four intensional concepts, those of convention, intention, legitimating reason and real reason. The first two let us identify what actors are doing, and the last two why they are doing it. The quartet is assembled as a final comment on Weber's demand for adequacy both at the level of meaning and at the causal level. The concluding balance between economic and social components of rational action is struck in chapter 12, where the Cunning of Reason is left as little work to do as possible. It is not wholly pensioned off, however, and I end somewhat messily, still holding out for a self which plays the roles, acts from the reasons and intervenes in the consequences.

Conclusion

Think of action as the result of desire plus belief and ask what marks out rational action. A promising answer is that a fully rational agent has fully ordered preferences, perfect information and faultless computing powers. He never misses an option with greater net expected utility. This is to take instrumental rationality (*Zweckrationalität*) as the primary and ideal-type case. Analysis sets off from microeconomics and decision-theory. It generalises from a single agent to several with the help of the theory of games and from economics to social life at large. With the ideal-type worked out, it can then be applied to actual life in Weberian spirit by using it, when it fits, to understand what happens or, when it does not, to identify what needs further explanation.

This ambitious programme runs into several troubles, some remediable, some not. It will not work even in its own terms

while its underlying account of human motivation is Humean. But a Kantian substitute restores its prospects. When it comes to generalising from one agent to several, however, a radical indeterminacy creeps in, because expectations are generative. There is often no one outcome which it is rational to expect, since the outcome depends on a regress of expectations. The indeterminacy is partly remedied, however, by there also being normative expectations. But this requires a different kind of agent, one who satisfices not just because of uncertainty but also because of economically incommensurable roles. Yet Adam is not the creature of his roles. He is a rational steward in office and, at the same time, retains a rational partial independence of the history which he helps to make. Singly and collectively he can intervene in the consequences of actions, if he is aware of the Cunning of Reason and its habit of making mischief. Contractarian thinking is thus a powerful tool of understanding, but even its especially forceful economic version implies that the social contract produces what Rousseau calls a remarkable change in man. Rational action is finally the expression of self in a social world.

That is the theme of the book in a stark form, which belies my respect for 'Modern Political Economy' or the attempt to generalise *homo economicus* across the board. So, to start us off, let me say that, if social science is to be possible, God cannot have given Adam a completely unconstrained hand to mould and fashion himself. There have to be some limitations on even a free and sovereign artificer. The theory of rational choice is an elegant, suggestive analysis of agent sovereignty in a constraining world.